

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

ENERO - JUNIO 2020

INGENIERÍA INFORMÁTICA

MATERIA

DATOS MASIVOS

CATEDRÁTICO:

JOSE CHRISTIAN ROMERO HERNANDEZ

UNIDAD 2

# PRACTICE 1

ALUMNO

López Valencia Luis Daniel #15212159

**we specify the work area**

setwd("C:/Users/Daniel/Desktop/Practicas R")

getwd()

**we choose our database to use**

dataset <- read.csv(file.choose())

dataset

**installing caTools library**

install.packages('caTools')

library(caTools)

**Splitting the dataset into the Training set and Test set**

set.seed(123)

split <- sample.split(dataset$comision, SplitRatio = 2/3)

training\_set <- subset(dataset, split == TRUE)

test\_set <- subset(dataset, split == FALSE)

**Fitting Simple Linear Regression to the Training set**

regressor = lm(formula = comision ~ Ventas,

data = dataset)

summary(regressor)

**Predicting the Test set results**

y\_pred = predict(regressor, newdata = test\_set)

**Visualising the Training set results**

library(ggplot2)

ggplot() +

geom\_point(aes(x=training\_set$Ventas, y=training\_set$comision),

color = 'red') +

geom\_line(aes(x = training\_set$Ventas, y = predict(regressor, newdata = training\_set)),

color = 'blue') +

ggtitle('ventas vs comision (Training Set)') +

xlab('Ventas') +

ylab('Comision')

**Visualising the Test set results**

ggplot() +

geom\_point(aes(x=test\_set$Ventas, y=test\_set$comision),

color = 'red') +

geom\_line(aes(x = training\_set$Ventas, y = predict(regressor, newdata = training\_set)),

color = 'blue') +

ggtitle('ventas vs comision (Test Set)') +

xlab('Ventas') +

ylab('Comision')